BAB II
TINJAUAN PUSTAKA

2.1 Landasan Teori

2.1.1 Aplikasi TikTok

Ada berbagai platform yang dapat digunakan masyarakat untuk berbagi informasi, dan salah satunya adalah TikTok. Aplikasi ini memungkinkan pengguna untuk membuat video singkat dan dinamis. Resmi diluncurkan pada tahun 2016 oleh Zhang Yiminy dari Cina, TikTok juga memacu kreativitas penggunanya dalam menciptakan konten video yang menarik. Bahkan, popularitas TikTok semakin meningkat, terutama selama pandemi, dengan peningkatan penggunaan hingga 20% dibandingkan dengan kondisi normal (Arofatul Hidayah dkk., 2024).

Tujuan utama TikTok adalah menyediakan platform yang menyenangkan dan kreatif bagi penggunanya untuk mengekspresikan diri mereka melalui video (Isnan dkk., 2023). Pengguna dapat membuat video yang mencakup berbagai konten seperti tarian, komedi, musik, dan banyak lagi. TikTok juga menawarkan berbagai fitur kreatif seperti filter, efek, dan musik latar yang dapat digunakan untuk mengedit video (Setiawan dkk., 2024).

Secara keseluruhan, TikTok adalah aplikasi yang menawarkan pengalaman unik dan menarik bagi penggunanya, memungkinkan mereka untuk mengekspresikan diri dan menemukan komunitas dengan minat yang sama di seluruh dunia (Rizal dkk., 2024).

2.1.2 NLP (*Natural Language Processing*)

Teknologi AI yang dikenal sebagai pemrosesan bahasa alami atau NLP,

memungkinkan mesin untuk memahami dan menganalisis bahasa manusia secara alami. NLP digunakan dalam konteks pencarian produk untuk membantu mesin dalam menguraikan makna dan maksud di balik istilah yang diketik pengguna (Setiawan dkk., 2024).

Pengaplikasian NLP sangat luas, mencakup pengolahan teks dan suara, klasifikasi teks, ekstraksi informasi, dan terjemahan Bahasa. Misalnya, *Grammarly* menggunakan NLP untuk memeriksa tata bahasa dan ejaan, sementara *Google Translate* menggunakan teknologi ini untuk menerjemahkan teks dari satu bahasa ke bahasa lain (Nur Oktavia dkk., 2024).

2.1.3 *Machine Learning*

Machine learning adalah cabang dari kecerdasan buatan (AI) yang memungkinkan sistem untuk belajar dan meningkatkan kinerjanya dari pengalaman tanpa harus diprogram secara eksplisit. Proses ini melibatkan penggunaan algoritma dan model statistik untuk menganalisis dan membuat prediksi berdasarkan data. Dalam *machine learning*, data adalah bahan baku utama yang digunakan untuk melatih model (Mohamad Sham & Mohamed, 2022). Data ini bisa berupa teks, gambar, suara, atau angka. Model adalah representasi matematis dari data yang digunakan untuk membuat prediksi atau keputusan berdasarkan data input. Proses pelatihan (*training*) adalah saat model belajar dari data dengan tujuan mengoptimalkan model untuk meminimalkan kesalahan prediksi. Setelah model dilatih, model diuji dengan data yang belum pernah dilihat sebelumnya untuk mengevaluasi kinerjanya, yang dikenal sebagai proses pengujian (*testing*) (Mustafa Taye, 2023).

Machine learning dibagi menjadi beberapa jenis, yaitu *supervised learning, unsupervised learning,* dan *reinforcement learning.* Dalam *supervised learning,* model dilatih dengan data yang sudah diberi label, seperti klasifikasi dan regresi. Dalam *unsupervised learning,* model dilatih dengan data yang tidak diberi label, seperti *clustering* dan asosiasi. Sedangkan dalam *reinforcement learning*, model belajar melalui trial and error untuk mencapai tujuan tertentu, seperti dalam permainan dan robotika (Christanto dkk., 2023).

Aplikasi *Machine learning* sangat luas dan mencakup berbagai bidang. Dalam computer vision, *Machine learning* digunakan untuk pengenalan wajah, deteksi objek, dan segmentasi gambar. Dalam *Natural Language Processing* (NLP), *Machine learning* digunakan untuk pengenalan suara, analisis sentimen, dan terjemahan bahasa (Mustafa Taye, 2023). Selain itu, *Machine learning* juga digunakan dalam sistem rekomendasi, seperti rekomendasi produk di *e-commerce* dan konten di *platform streaming*. Dengan kemampuannya untuk belajar dan beradaptasi dari data, *Machine learning* memiliki potensi besar untuk mengubah berbagai industri dan aspek kehidupan kita (Mohamad Sham & Mohamed, 2022).

2.1.4 Analisis Sentimen

Analisis sentimen adalah tindakan menentukan bagaimana perasaan diekspresikan dalam teks dengan memeriksa data tekstual dan menguraikan sentimennya. Melalui pemrosesan informasi, analisis sentimen dapat digunakan untuk memeriksa sudut pandang yang muncul dalam evaluasi pengguna di *Google Play Store* (Oktavia dkk., 2023). Pengambilan review dilakukan dengan metode *Scraping*, mengekstraksi, dan memproses data teks secara mandiri untuk

menghasilkan data efektif yang mengindikasikan apakah opini tersebut positif, negatif atau netral. Analisis sentimen diperlukan karena kesulitan dalam memahami dan bereaksi terhadap ulasan pengguna. Memanfaatkan situs web seperti TikTok menghasilkan banyak data (Fathoni Naufal Fernanda dkk., 2024). Tantangan untuk mengatur dan memahami evaluasi dalam jumlah besar secara manual dapat menimbulkan masalah. Dalam mengatasi tantangan ini, dapat melakukan solusi otomatisasi dan teknologi berbasis kecerdasan buatan (AI). Misalnya, teknik *Natural Language Processing* (NLP) dapat digunakan untuk menganalisis teks evaluasi secara otomatis, mengidentifikasi pola dan tren yang mungkin tidak terlihat oleh manusia. Algoritma *Machine learning* juga dapat dilatih untuk memberikan penilaian yang lebih konsisten dan objektif (Nur Oktavia dkk., 2024).

2.1.5 *Labelling*

Labelling merupakan proses memberikan label atau kategori tertentu pada data berdasarkan karakteristik atau isi dari data. Dalam analisis sentiment, *labelling* digunakan untuk menentukan apakah suatu teks memiliki sentimen positif, negatif, atau netral (Asri et al., 2022).

Cara kerja *labelling* akan melakukan pengumpulan data mentah, seperti teks ulasan, komentar, atau postingan. Selanjutnya, setiap data akan dianalisis untuk menentukan makna atau konteksnya, lalu diberi label yang sesuai. Pada penelitian ini akan menggunakan *VADER Lexicon* dan *TextBlob* (Ayuningtyas dkk., 2024).

TextBlob adalah pustaka Python yang menyediakan alat untuk penambangan teks, analisis teks, dan pemrosesan teks. Metode yang digunakan

oleh *TextBlob* adalah berbasis *lexicion*. *TextBlob* menggunakan daftar fitur leksikal (misalnya kata-kata) yang diberi label sebagai positif atau negatif berdasarkan orientasi semantiknya (Harris Pratama & Hayaty, 2023). *TextBlob* menggunakan *Lexicon – based approach* yang mana setiap kata memiliki skor polaritas yang menunjukkan tingkat positif atau negatif. Hasilnya memiliki skor polaritas kata : Nilai antara -1 hingga 1, di mana -1 menunjukkan sentimen negatif dan +1 menunjukkan sentimen positif, *TextBlob* juga menghitung *subjectivity* 0 (Fakta) hingga 1 (opini) (Sid, 2022).

Rumus perhitungan Polarity

$$Polarity_{kalimat} = \frac{\sum_{i=1}^{n} Polarity_{katai}}{n} \quad 2.1$$

n = jumlah kata yang dikenali dalam *lexicon*.

Hasil:

Polarity > 0 → Positif

Polarity < 0 → negatif

Polarity = 0 → Netral

VADER Lexicon juga merupakan pustaka analisis sentimen berbasis *lexicon* dengan aturan yang telah ditentukan untuk kata-kata atau *lexicon, VADER Lexicon* untuk teks sosial media dan informal . Yang membedakan *VADER Lexicon* adalah kemampuannya untuk tidak hanya mengklasifikasikan kata-kata sebagai positif, negatif, atau netral, tetapi juga mengevaluasi sentimen keseluruhan dari sebuah kalimat. *VADER Lexicon* mengembalikan probabilitas kalimat input sebagai positif, negatif, dan netral (Putri Meliani dkk., 2022a).

Setiap kata diberi ***valence score*** (-4 hingga +4)

Memperhitungkan :

Negasi ⟶ membalik skor kata jika ada kata seperti *not* atau *never*.

Intensifier/Booster ⟶ kata seperti *very, extremely* menambah intensitas.

***Punctuation* dan Emoji** ⟶ tanda seru (!) atau emoji mempengaruhi skor.

Conjunction ⟶ kata seperti *but* memberi bobot lebih pada kata setelahnya.

Rumus Perhitungan *Compound Score*

$$compound = \frac{\sum_{i=1}^{n} valence_i}{\sqrt{\sum_{i=1}^{n} valence_i^2 + \alpha}} \quad 2.2$$

α biasanya = 15 (untuk normalisasi skor ke-1 sampai +1)

Interpretasi skor:

Compound ≥ 0.05 → Positif

Compound ≤ -0.05 → Negatif

-0.05 < *Compound* < 0.05 → Netral

2.1.6 Xgboost (*eXtreme Gradient Boosting*)

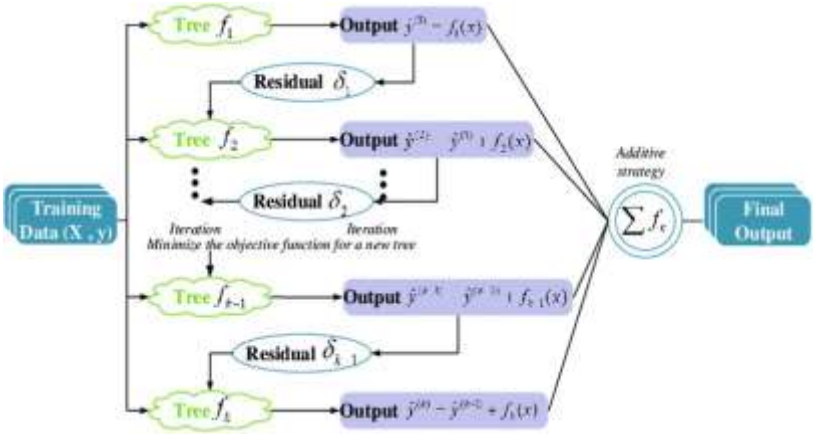
eXtreme Gradient Boosting (XGBoost) sebuah algoritma pembelajaran mesin yang merupakan pengembangan dari algoritma *gradient boosting*. *XGBoost* menggunakan teknik pembelajaran *ensemble*, di mana beberapa model digabungkan untuk meningkatkan akurasi prediksi. Pada dasarnya, *XGBoost* menggunakan metode pohon keputusan (*decision trees*) sebagai pembelajar dasar (*base learner*), dimana setiap pohon dibangun secara iteratif dengan menambahkan satu pohon pada satu waktu ke dalam model (Maulana dkk., 2023).

Pada setiap iterasi, *XGBoost* menghitung kesalahan residual dari model sebelumnya dan berusaha untuk memperbaiki prediksi pada iterasi berikutnya. Dalam upaya meningkatkan kualitas model, *XGBoost* menerapkan teknik

regularisasi dan pemangkasan (*pruning*) (Thongsuwan dkk., 2021). Regularisasi berfungsi untuk mengontrol kompleksitas model dan menghindari overfitting, yaitu kondisi di mana model terlalu disesuaikan dengan data latih sehingga performanya menurun pada data uji. Sementara itu, pemangkasan bertujuan untuk menghindari terbentuknya cabang-cabang pohon yang tidak signifikan dalam memprediksi variabel target (Angga Purnama Widiarta dkk., 2023).

Selain itu, *XGBoost* juga menggunakan metode *gradient descent* dalam proses penyesuaian (optimasi) parameter model. Algoritma ini menghitung nilai gradien dari fungsi loss pada setiap iterasi dan menggerakkan parameter model ke arah yang mengurangi nilai loss tersebut. Beberapa keunggulan *XGBoost* meliputi kemampuannya dalam menangani data berukuran besar dengan efisien, kemampuan untuk mengestimasi pentingnya fitur-fitur, serta tingkat akurasi prediksi yang tinggi. Namun, dalam penggunaannya, terdapat beberapa hal yang perlu diperhatikan, seperti penyetelan hyperparameter agar model menghasilkan prediksi yang optimal, mempertimbangkan kemungkinan terjadinya overfitting, dan memperhatikan interpretasi dari hasil prediksi yang dihasilkan oleh model (Natras dkk., 2022).

Dengan memahami mekanisme dan karakteristik *XGBoost*, pengguna dapat memanfaatkan algoritma ini secara efektif dalam berbagai aplikasi pembelajaran mesin, terutama ketika dihadapkan pada tantangan data yang kompleks dan berukuran besar. Berikut disajikan gambar berupa arsitektur algoritma *XGBoost* pada Gambar 2.1.



Gambar 2.1 Arsitektur Algoritma *XGBoost* (Aqbar & Andrianti Supono, 2023)

2.1.7 Pengujian Model

Confusion matrix adalah salah satu cara yang paling umum digunakan untuk mengevaluasi kinerja model klasifikasi, termasuk model analisis sentimen seperti *VADER Lexicon* dan *TextBlob*. *Confusion matrix* memberikan gambaran tentang bagaimana prediksi model sesuai dengan label sebenarnya dalam bentuk matriks, yang memungkinkan kita untuk menghitung berbagai metrik kinerja seperti *akurasi*, *presisi*, *recall*, dan *F1-score*.

Confusion Matrix terdiri dari empat komponen utama yang membentuk tabel 2x2:

***True Positive* (TP)**	:	Jumlah data yang benar-benar positif dan diprediksi positif oleh model.
***True Negative* (TN)**	:	Jumlah data yang benar-benar negatif dan diprediksi negatif oleh model.
***False Positive* (FP)**	:	Jumlah data yang benar-benar negatif tetapi diprediksi positif oleh model. Kesalahan ini juga :dikenal sebagai Type I Error.

***False Negative* (FN)** Jumlah data yang benar-benar positif tetapi di prediksi negatif oleh model. Kesalahan ini juga dikenal sebagai Type II Error.

Tabel 2. 1 *Confusion Matrix*

	Predicted Positive	***Predicted Negative***
Actual Positive	*True Positive (TP)*	*False Negative (FN)*
Actual Negative	*False Positive (FP)*	*True Negative (TN)*

Rumus-rumus penting dalam *Confusion Matrix :*

***Accuracy* (Akurasi) :** Mengukur proporsi prediksi yang benar terhadap seluruh prediksi.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3)$$

Akurasi tinggi menunjukkan bahwa model memberikan prediksi yang benar untuk sebagian besar data.

***Precision* (Presisi) :** Mengukur proporsi prediksi positif yang benar terhadap semua prediksi positif.

$$Precision = \frac{TP}{TP + FP} \quad (2.4)$$

Presisi tinggi menunjukkan bahwa sebagian besar data yang diprediksi positif memang benar-benar positif.

Recall (Sensitivity/Recall) : Mengukur proporsi data positif yang benar-benar diprediksi positif oleh model.

$$Recall = \frac{TP}{TP + FN} \quad (2.5)$$

Recall tinggi menunjukkan bahwa model mampu mendeteksi sebagian besar data positif.

F1-Score : Menggabungkan *presisi* dan *recall* dalam satu metrik harmonis. F1-Score berguna ketika terdapat ketidakseimbangan kelas.

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} \times \text{Recall}} \quad 2.6$$

Specificity : Mengukur proporsi data negatif yang benar-benar diprediksi negatif oleh model.

$$Specificity = \frac{TN}{TN + FP} \quad 2.7$$

2.2 Penelitian Terkait (*State-Of-The-Art*)

Tabel 2.2 *State Of The Art*

No	Peneliti	Judul	Tujuan	Persamaan Penelitian Terkait dengan Penelitian Yang Dilakukan	Perbedaan Penelitian Terkait dengan Penelitian yang Dilakukan
1.	(Siva Asvia, 2023)	Perbandingan *Naïve Bayes* dan *Random forest* pada Analisis Sentimen Reputasi Brand Provider Live.on	Membandingkan performa algoritma *Naïve Bayes* dan *Random forest* pada hasil analisis sentiment *Provider Live.on* dari komentar di Google Play	Membahas Analisis Sentimen Mengenai suatu aplikasi tertentu.	**Penelitian Terkait:** Membandingkan *Naïve Bayes* dan Random Forest, Fokus pada Analisis Reputasi Brand Provider Live.On. **Penelitian yang Dilakukan:** Menggunakan algoritma XGBoost Fokus pada Perbandingan labelling.
2.	(Yuda Maulana, 2023)	Optimalisasi *Support Vector Machine* (SVM) Menggunakan Pelabelan Vader Pada	Menganalisis performa algoritma *Support Vector Machine* (SVM) dalam melakukan analisis	Menggunakan teknik pelabellan pada analisis sentimen.	**Penelitian Terkait:** menggunakan metode pelabelan VADER untuk memberikan label sentimen. **Penelitian yang Dilakukan:**

No	Peneliti	Judul	Tujuan	Persamaan Penelitian Terkait dengan Penelitian Yang Dilakukan	Perbedaan Penelitian Terkait dengan Penelitian yang Dilakukan
		Analisis Sentimen Ulasan *Google Classroom*	sentimen pada ulasan Google Classroom dengan menerapkan metode pelabelan *VADER*		Menggunakan perbandingan teknik *VADER lexicon* dan *Textblob*.
3.	(Ezra Rofran & Joanda Kaunang , 2024)	Analisisis Sentimen Pengguna Instagram Terhadap Kebijakan Nadiem Makarim yang Memperbolehkan mahasiswa lulus tanpa skripsi menggunkan metode analisis vader dan metode klasifikasi *Naïve Bayes*.	Membandingkan hasil pelabelan menggunakan *VADER* dengan metode Crowdsourcing.	Membahas Analisis Sentimen Mengenai suatu aplikasi tertentu.	**Penelitian Terkait:** Menggunakan Metode analisis VADER untuk pelabellan sentimen dan algoritma *Naïve Bayes* untuk klasifikasi sentiment. **Penelitian Yang Dilakukan:** Melakukan perbandingan pelabelan antara *VADER lexicon* dan *textblob*
4.	(Safitri dkk.,	Perbandingan algoritma *XGBoost* dan SVM	Membandingkan efektivitas *XGBoost* dan	Melakukan Analisis Sentimen	**Penelitian Terkait:** Perbandingan kinerja dua algoritma klasifikasi,

No	Peneliti	Judul	Tujuan	Persamaan Penelitian Terkait dengan Penelitian Yang Dilakukan	Perbedaan Penelitian Terkait dengan Penelitian yang Dilakukan
	2024)	dalam analisis opini publik pemilihan presiden 2024	SVM dalam mengklasifikasikan opini public di Twitter mengenai calon presiden.	Menggunakan algoritma XGBoost.	yaitu *Xgboost* dan SVM dalam analisis opini publik terkait pemilihan presiden indonesia 2024. **Penelitian Yang Dilakukan:** Membandingkan proses labelling *VADER lexicon* dan *Textblob*.
5.	(Oktavia dkk., 2023)	Analisis Sentimen Terhadap Penerapan sistem E-Tilang Pada Media Sosial Twitter Menggunakan algoritma *Support Vector Machine* (SVM)	Menganalisis sentiment pengguna Twitter terhadap sistem e- Tilang untuk mengidentifikasi opini positif, negatif atau netral.	Menggunakan proses analisis setimen pada suatu aplikasi.	**Penelitian Terkait:** Analisis sentimen terhadap penerapan sistem e-Tilang di media social Twitter menggunakan algoritma *Support Vector Machine* (SVM). **Penelitian Yang Dilakukan:** Analisis Sentimen Aplikasi

No	Peneliti	Judul	Tujuan	Persamaan Penelitian Terkait dengan Penelitian Yang Dilakukan	Perbedaan Penelitian Terkait dengan Penelitian yang Dilakukan
					Tiktok menggunakan Algoritma Xgboost
6.	(Husdi & Kamaruddin, 2024)	Auto Labelling untuk Analisis sentiment Opini Mahasiswa Baru Terhadap Pembelajaran Mata Kuliah dengan Algoritma *K- Nearest Neighbor*	Menganalisis sentimen mahasiswa baru untuk mengevaluasi dan memperbaiki layanan akademik.	Melakukan proses labelling pada analisis sentimen.	**Penelitian Terkait:** Analisis sentiment opini mahasiswa baru terhadap pembelajaran mata kuliah di Fakultas Ilmu Komputer Universitas Ichsan Gorontalo. **Penelitian Yang Dilakukan:** Melakukan analisis sentimen pada review aplikasi TikTok.
7.	(Dava Maulana dkk., 2023)	Algoritma *Xgboost* Untuk Klasifikasi Kualitas Air Minum	Mengukur akurasi algoritma *XGBoost* dalam mengklasifikasikan air minum yang layak dan tidak layak.	Menggunakan Algoritma *XGBoost*.	**Penelitian Terkait:** algoritma *XGBoost* untuk klasifikasi kualitas air minum. **Penelitian Yang Dilakukan:** Penggunaan algoritma *XGBoost* untuk proses analisis sentimen.

No	Peneliti	Judul	Tujuan	Persamaan Penelitian Terkait dengan Penelitian Yang Dilakukan	Perbedaan Penelitian Terkait dengan Penelitian yang Dilakukan
8.	(Fazrin dkk., 2023)	Perbandingan Algoritma *K- Nearest Neighbor* dan *Logistic Regression* pada Analisis Sentimen terhadap Vaksinasi Covid-19 pada Media Sosial Twitter dengan Pelabelan VADER dan *Textblob*	Mengukur tingkat akurasi kedua algoritma dengan pelabelan menggunakan *TextBlob* dan *Vader Sentiment*.	Menggunakan teknik pelabelan vader dan Textblob pada analisis sentimen.	**Penelitian Terkait:** perbandingan algoritma *K- Nearest Neighbor* (KNN) dan Logistic Regression dalam analisis sentimen terhadap vaksinasi Covid- 19 di media sosial Twitter. **Penelitian Yang Dilakukan:** Melakukan perbandingan terhadap teknik *labelling* pada algoritma XGBoost.
9.	(Christanto dkk., 2023)	Analisis Perbandingan *Decision Tree, Support Vector Machine*, dan XGBoost dalam Mengklasifikasi Review Hotel Trip Advisor	menentukan algoritma terbaik yang dapat digunakan untuk analisis sentimen review hotel, sehingga memudahkan traveler dalam memilih	Menggunakan algoritma *XGBoost*.	**Penelitian Terkait:** menganalisis dan membandingkan tiga algoritma *Machine Learning*, yaitu *Decision Tree*, *Support Vector Machine* (SVM), dan XGBoost, dalam mengklasifikasikan review

No	Peneliti	Judul	Tujuan	Persamaan Penelitian Terkait dengan Penelitian Yang Dilakukan	Perbedaan Penelitian Terkait dengan Penelitian yang Dilakukan
			hotel berdasarkan review pengguna sebelumnya.		hotel di platform Trip Advisor **Penelitian Yang Dilakukan:** Menggunakan algoritma XGBoost untuk analisis sentimen.
10.	(Harris Pratama & Hayaty, 2023a)	*Performance of Lexical Resource and Manual Labeling on Long Short-Term Memory Model for Text Classification*	Membandingkan hasil pelabelan manual dengan pelabelan menggunakan sumber daya lexicon.	Melakukan teknik *labelling*.	**Penelitian Terkait:** Menggunakan Algoritma LSTM. **Penelitian Yang Dilakukan:** Menggunakan algoritma XGBoost.
11.	(Rashid dkk., 2024)	*A Multilingual Corpus for Panic and Worry in Code- Mixed Tweets by VADER Sentiment Analysis*	Untuk mengidentifikasi ekspresi kecemasan dalam berbagai konteks linguistik	Menggunakan teknik pelabelan pada suatu hal.	**Penelitian Terkait:** Menggunakan analisis sentimen berbasis korpus dengan VADER. **Penelitian Yang Dilakukan:** Menggunakan perbandingan teknik pelabelan *VADER lexicon*

No	Peneliti	Judul	Tujuan	Persamaan Penelitian Terkait dengan Penelitian Yang Dilakukan	Perbedaan Penelitian Terkait dengan Penelitian yang Dilakukan
					dan *Textblob*.
12.	(Sugant hini dkk., 2023)	*Sentiment Analysis on Multilingual Tweets using XGBoost Classifiers*	Menggunakan analisis semantik sebagai fitur tambahan untuk meningkatkan akurasi model klasifikasi	Menggunakan algoritma XGBoost pada analisis sentiment.	**Penelitian Terkait:** Mengumpulkan tweet dalam berbagai bahasa dan mengklasifikasikannya menjadi tweet positif dan negatif **Penelitian Yang Dilakukan:** Mengumpulkan dataset dari review *google play store* pada TikTok.
13.	(Huang & Wang, 2023)	*Impact of COVID- 19 on mental health in China: analysis based on sentiment knowledge enhanced pre- training and XGBoost algorithm*	Menganalisis dampak COVID- 19 terhadap kesehatan mental masyarakat di Tiongkok menggunakan data teks sosial dari platform Sina Weibo	Melakukan proses analisis sentimen.	**Penelitian Terkait:** menggunakan teknologi pemrosesan bahasa alami (NLP) dan algoritma XGBoost. **Penelitian Yang Dilakukan:** Melakukan perbandingan teknik Labelling pada *VADER Lexicon* dan *Textblob*.

No	Peneliti	Judul	Tujuan	Persamaan Penelitian Terkait dengan Penelitian Yang Dilakukan	Perbedaan Penelitian Terkait dengan Penelitian yang Dilakukan
14.	(Fathoni Naufal Fernanda dkk., 2024)	Perbandingan Performa *Labeling Lexicon* InSet dan *VADER* pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM	Membandingkan performa dua kamus *lexicon*, yaitu InSet dan *VADER*	Melakukan perbandingan teknik labelling.	**Penelitian Terkait:** analisis sentimen terkait Rohingya di media sosial X menggunakan metode *Support Vector Machine* (SVM) **Penelitian Yang Dilakukan:** Menggunakan algoritma XGBoost.

Pada tabel 2.1 merupakan rangkuman beberapa penelitian yang fokus pada analisis sentimen menggunakan berbagai algoritma dan metode pelabelan. Setiap kolom menggambarkan penelitian terkait dengan topik analisis sentimen, algoritma yang digunakan, serta teknik pelabelan yang diterapkan. Mayoritas penelitian membahas analisis sentimen dari data media sosial atau platform tertentu (seperti *Google Play*, *Twitter*, dan *Instagram*), serta beragam aplikasi seperti e-Tilang, vaksinasi, kualitas air minum, dan pendidikan. Algoritma yang sering dibahas mencakup *Naïve Bayes*, *Random Forest*, *Support Vector Machine* (SVM), XGBoost, *K-Nearest Neighbor* (KNN), *Logistic Regression*, dan *Long Short-Term Memory* (LSTM). Algoritma ini digunakan untuk mengklasifikasikan opini atau sentimen menjadi kategori seperti positif, negatif, atau netral. Banyak penelitian membandingkan

teknik pelabelan sentimen menggunakan metode seperti *VADER*, *TextBlob*, dan InSet, yang bertujuan untuk memberikan label sentimen secara otomatis. Penelitian-penelitian ini memiliki kesamaan dalam hal penggunaan algoritma untuk mengklasifikasikan sentimen dari data teks serta penerapan teknik pelabelan otomatis untuk mengidentifikasi opini publik. Beberapa penelitian membahas aplikasi atau domain yang berbeda (seperti analisis reputasi brand, kualitas air, kebijakan pendidikan, hingga kesehatan mental). Selain itu, terdapat perbedaan dalam kombinasi algoritma dan teknik pelabelan yang digunakan, seperti XGBoost vs SVM, atau *VADER* vs *TextBlob*.

2.3 Matriks Penelitian

Tabel 2. 3 Matriks Penelitian

No	Judul	Penulis	Ruang Lingkup									
			Teknik pelabelan			Algoritma						
			VADER	VADER Lexicon	Textblob	Svm	Naïve Bayes	Random Forest	KNN	XGBoost	Decision Tree	LSTM
1.	Perbandingan *Naïve Bayes* dan *Random forest* pada Analisis Sentimen Reputasi Brand Provider Live.on	(Siva Asvia, 2023)					√	√				
2.	Optimalisasi *Support Vector Machine* (SVM) Menggunakan Pelabelan Vader Pada Analisis Sentimen Ulasan *Google Classroom*	(Yuda Maulana, 2023)	√			√						
3.	Analisisis Sentimen Pengguna Instagram Terhadap Kebijakan Nadiem Makarim yang Memperbolehkan mahasiswa lulus tanpa skripsi menggunkan metode analisis vader dan metode klasifikasi *Naïve Bayes*	(Ezra Rofran & Joanda Kaunang, 2024)		√			√					

No	Judul	Penulis	Ruang Lingkup									
			Teknik pelabelan			Algoritma						
			VADER	VADER Lexicon	Textblob	Svm	Naïve Bayes	Random Forest	KNN	XGBoost	Decision Tree	LSTM
4.	Perbandingan algoritma XGBoost dan SVM dalam analisis opini publik pemilihan presiden 2024	(Safitri dkk., 2024)				√				√		
5.	Analisis Sentimen Terhadap Penerapan sistem E-Tilang Pada Media Sosial *Twitter* Menggunakan algoritma *Support Vector Machine* (SVM)	(Oktavia dkk., 2023)				√						
6.	Auto Labelling untuk Analisis sentiment Opini Mahasiswa Baru terhadap Pembelajaran Mata Kuliah dengan Algoritma *K-Nearest Neighbor*	(Husdi & Kamaruddin, 2024)							√			
7.	Algoritma Xgboost Untuk Klasifikasi Kualitas Air Minum Perbandingan *Algoritma K- Nearest Neighbor* dan *Logistic*	(Dava Maulana dkk., 2023)								√		
8.	Regression pada Analisis Sentimen terhadap Vaksinasi Covid- 19 pada Media Sosial *Twitter*	(Fazrin dkk., 2023)		√	√				√			

No	Judul	Penulis	Ruang Lingkup									
			Teknik pelabelan			Algoritma						
			VADER	VADER Lexicon	Textblob	Svm	Naïve Bayes	Random Forest	KNN	XGBoost	Decision Tree	LSTM
	dengan Pelabelan *Vader* dan *Textblob*											
9.	Analisis Perbandingan *Decision Tree, Support Vector Machine*, dan Xgboost dalam Mengklasifikasi Review Hotel Trip Advisor	(Safitri dkk., 2024)				√				√	√	
10	*Performance of Lexical Resource and Manual Labeling on Long Short-Term Memory Model for Text Classification*	(Harris Pratama & Hayaty, 2023b)		√								√
11	*A Multilingual Corpus for Panic and Worry in Code-Mixed Tweets by VADER Sentiment Analysis*	(Rashid dkk., 2024)	√			√		√				
12	*Sentiment Analysis on Multilingual Tweets using XGBoost Classifiers*	(Suganthini dkk., 2023)			√					√		
13.	*Impact of COVID-19 on mental health in China: analysis based on sentiment knowledge*	(Huang & Wang, 2023)								√		

No	Judul	Penulis	Ruang Lingkup									
			Teknik pelabelan			Algoritma						
			VADER	VADER Lexicon	Textblob	Svm	Naïve Bayes	Random Forest	KNN	XGBoost	Decision Tree	LSTM
	enhanced pre- training and XGBoost algorithm											
14.	Perbandingan Performa Labeling *Lexicon InSet* dan VADER pada Analisa Sentimen Rohingya di Aplikasi X dengan SVM	(Fathoni Naufal Fernanda dkk., 2024)	√	√		√						